A Review - Secured Approach for Hybrid Cloud De-Duplication

Kakkad Komal, Bagwan Karishma, Shaikh Sumayya, Prof. Murkute. P. K., Prof. Naved Raza Q. Ali.

Al-Ameen College of Engineering,Koregaon Bhima,Savitribai Phule Pune University, Pune, India

ABSTRACT

A hybrid cloud is a combination of public and private clouds bound together by either standardized or proprietary technology that enables data and application portability. Proposed system aiming to efficiently solving the problem of deduplication with differential privileges in cloud computing. A hybrid cloud architecture consisting of a public cloud and a private cloud and the data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud. To make data management scalable in cloud computing, deduplication has been a very well-known technique recently is use. Deduplication reduces your bandwidth requirements, speeds up the data transfers, and it keeps your cloud storage needs to a minimum. Proposed system present several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture. To maintain the confidentiality of data the convergent encryption technique has been used to encrypt the data before outsourcing. Authorized deduplication system support differential authorization duplicate check. As a proof of concept, a prototype is implemented in authorized duplicate check scheme and conduct test bed experiments using prototype, authorized duplicate check scheme incurs minimal overhead compared to normal operations.

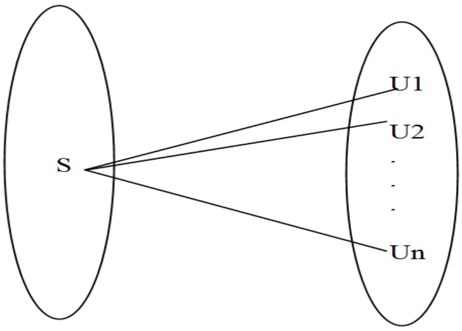
Keywords: Deduplication, Authorized Duplicatecheck, Hybrid Cloud, Covergent Encryption.

I. INTRODUCTION

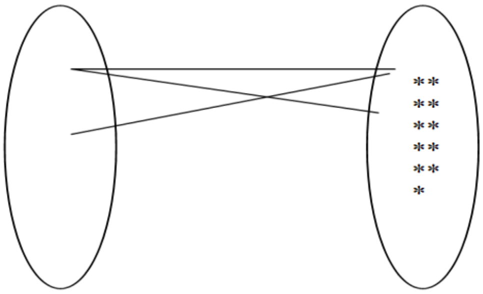
Relevant Mathematics Associated with the Dissertation:

1 Mapping

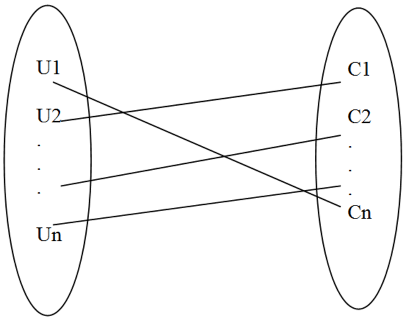
One system will have many users. Hence one too many relationship is observed here.



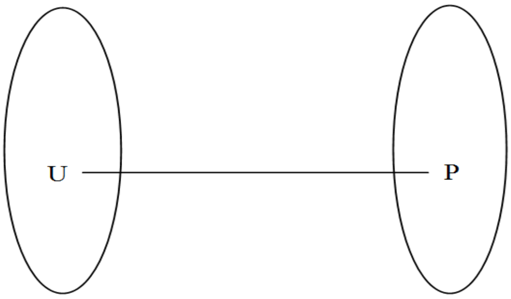
Ignore this one.



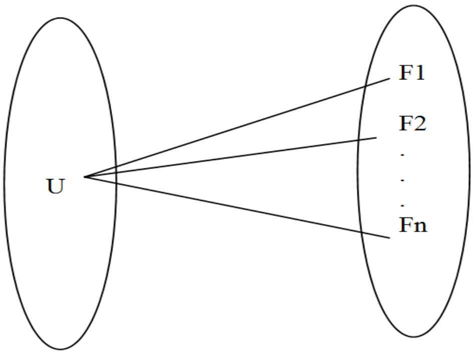
Multiple upload requests of same file can occur from many users, thus. Hence many to many relationship is observed here.



One user will have his/her own private Cloud. Hence one to one relationship will be observed here.



One user can upload many files. Hence one to many relationship is observed here.



2. Set Theory

Our system can be represented as a set

X = {I, O, S_C, F_C, C}

where,

I=set of inputs

O=set of outputs

S_C= set of outputs in success cases

F_C = set of outputs in failure cases

C = set of constraints

I = {U_D, F_R, T_R, K}

Where,

U_D = set of user details

F_R = set of requests by users for file upload

T_R = set of requests by users for token

K = set of encryption/decryption keys

O = {F_D, F_U, F_M}

where,

F_D = set of files downloaded

F_U = set of files uploaded

F_M = set of files modified

S_C = {F_{Dn}, F_{Un}, F_{Mn}}

where,

F_{Dn} = valid set of files downloaded

F_{Un} = valid set of files uploaded

F_{Mn} = valid set of files modified

F_C = {F_{Dn}, F_{Un}, F_{Mn}, NULL}

where,

F_{Do} = invalid set of files downloaded

F_{Uo} = invalid set of files uploaded

F_{Mo} = invalid set of files modified

NULL represents no output

C = {C_1, C_2, C_3}

where,

C_1 = “Every file must be encrypted before uploading a file into the cloud”

C_2 = “All users must be authenticated”

C_3 = “All files must be decrypted after downloading from the cloud”

U_D = {U_{D1}, U_{D2},…,U_{Dn}}

where,

U_{D1}, U_{D2}, …,U_{Dn} are details of the user.

T_R = {T_{R1}, T_{R2},…,T_{Rn}}

where,

T_{R1}, T_{R2}, …,T_{Rn} are token requests from users.

K = {K_1, K_2,…,K_n}

where,

K_1, K_2,…,K_n are encryption/decryption keys.

F_D, F_{Do}, F_{Dn}, F_U, F_{Uo}, F_{Un}, F_M, F_{Mo}, F_{Mn} are in the form

F = {F_1, F_2,…,F_n}

where,

F_1, F_2,…,F_n are files.

II. METHODS AND MATERIAL

3. Goal of Project

It excludes the security problems that may arise in the practical deployment of the present model. Also, it increases the national security. It saves the memory by deduplicating the data and thus provide us with sufficient memory. It provides authorization to the private firms and protects the confidentiality of the important data.

4. Literature Survey

In [1] C.-K Huang, L.-F Chien, and Y.-J Oyang, Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query Session Logs "2013. proposes a best method for basic encryption is transformed from a single file to a chunk. The symmetric key is used to cipher text formation. These keys are generated by the chunk file content.

In [4] Ian Clarke, Oskar Sandberg, Brandon Wiley, and Theodore W. Hong " A secure cloud backup system with assured deletion and version control ", 2013. mentioned the peer-to-peer network application which store data effectively to the database and retrieval makes easy.

In [5] Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou" Secure deduplication with efficient and reliable convergent key management ", 2013.Mention the efficient and very much reliable method of data de-duplication by using conversion encryption techniques. here theres is secure distribution of key management these convergent keys shares across multiple server.

In [7] John Douceur, Atul Adya, William J. Bolosky, Dan Simon and Marvel Theimer "Reclaiming space from duplicate files in a serverless distributed file system",2013.Make focus on distributed computing. To utlise the memory they used the convergent encryption technique in which the identical files are encrypted with different method still the different file is get stored.

In [9] K. Zhang, X. Zhou, Y. Chen, X. Wang, Y. Ruan" privacy ware data intensive computing on hybrid clouds," J. Am. Soc. for Information science and Technology, vol. 54, no. 7, 2011. proposes a hybrid cloud approach that support the data framework. These hybrid cloud introduced the public cloud and the private cloud . the sensitive information computation done by private cloud and non-sensitive information is present in the public cloud.

In [11] M. Bellare, S. Keelveedhi, and T. Ristenpart "Message-locked encryption and secure deduplication ", 2014. Mainly concerned with distinction between de-duplication and encryption. De-duplication can save only single file in case of encryption. Same file and same data decoded with two different keys result with different cipher text. That means same file and same content will be stored in the server. To overcome this problem a new approach introduced in which encryption keys are generated in consistent approach from block of data. there is two approach for data de-duplication, authenticated and anonymous .both method can be introduced to single server as well as distributed storage server approach. In this approach we used the convergent encryption technique and also in this approach plain text is also never sending to the server all the decode operation is done at the clientside.

In [12] M. Bellare and A. Palacio. Gq and schnorr identification schemes: "Proofs of security against impersonation under active and concurrent attacks" , 2013. mainly gives the outsourcing of the data . cloud computing gives more simple and effective approach for data storage. when data is outsource many security related issues and problems are comes inti picture such as malicious code running on the cloud. There may be a leakage of data. the primary requirement of cloud client is the confidentiality of the data. In proposed model the author introduced the twin cloud and commodity cloud. The working of the two cloud is different. The authorized cloud performs security operation such as encryption and commodity cloud perform time critical operation on the encrypted data. the client first of all make interaction with authorized cloud and then commodity cloud.

In [13] M. Bellare, C. Namprempre, and G. Neven "Security proofs for identity-based identification and signature schemes", 2012. Here the client has to prove his identity by proof of ownership protocol. a client who keep the private data proves the server thart he or she keeping the summery string of the file.

5. Existing System

Data deduplication systems, the private cloud is involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges.

Such architecture is practical and has attracted much attention from researchers.

The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud.

6. Aim

The aim of this project is to develop a system to address the problem of privacy preserving deduplication in cloud computing and propose a new deduplication system supporting for :

1. Differential Authorization
2. Authorized Duplicate Check
3. Unforgeability of file token/duplicate-check token
4. Indistinguishability of file token/duplicate-check token
5. Data Confidentiality

III. RESULTS AND DISCUSSION

7. Proposed System

In our system we will be implementing a project that includes the public cloud and the private cloud and also the hybrid cloud which is a combination of the both public cloud and private cloud. In general by if we used the public cloud we can’t provide the security to our private data and hence our private data will be loss. So that we have to provide the security to our data for that we make a use of private cloud also.

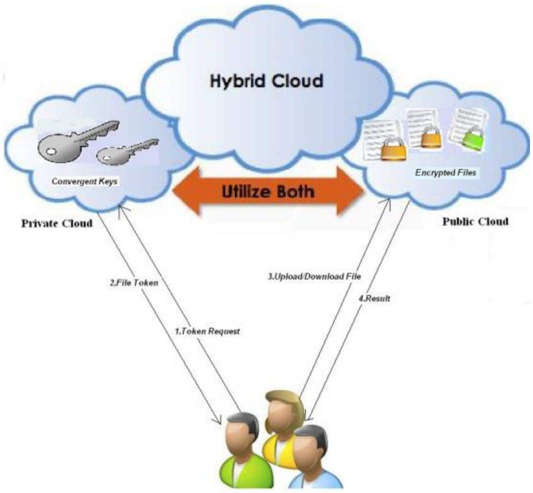


Figure 1. Architecture of Authorized Deduplication

When we use private clouds the greater security can be provided. In this system we also provides the data deduplication . Which is used to avoid the duplicate copies of data .User can upload and download the files from public cloud but private cloud provides the security for that data .that means only the authorized person can upload and download the files from the public cloud.

Data User

A user is an entity that want to access the data or files from S-SCP. User generate the key and store that key in private cloud. In storage system supporting deduplication, The user only upload unique data but do not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users. Each file is protected by convergent encryption key and can access by only authorized person. In our system user must need to register in private cloud for storing token with respective file which are store on public cloud. When he want to access that file he access respective token from private cloud and then access his files from public cloud. token consist of file content F and convergent key KF.

Private Cloud

In general for providing more security user can use the private cloud instead of public cloud. User store the generated key in private cloud. At the time of downloading system ask the key to download the file. User cannot store the secrete key internally. for providing proper file. When user want to access the key he first check authority of user then an then provide key. Protection to key we use private cloud. Private cloud only store the convergent key with respective.

Public Cloud:

Public cloud entity is used for the storage purpose. User uploads the files in public cloud. Public cloud is similar as S-CSP. When the user wants to download the files from public cloud, it will be ask the key which is generated or stored in private cloud. When the users key is match with files key at that time user can download the file, without key user cannot access the file. Only authorized user can access the file. In public cloud all files are stored in encrypted format. If any chance unauthorized person hack our file, but without the secrete or convergent key he doesn’t access original file. On public cloud there are lots of files are store each user access its respective file if its token matches with S-CSP server token.

8. Applications:

Hybrid clouds are mainly built to suit any of the IT environment or architecture, whether it might be any enterprise wide IT network or any department. Public data which is stored can be analyzed from statistical analyses which is done by social media, government entities can be used to enhance and analyze their own corporate data stand which is internal to gain the most form of perusing hybrid cloud Benefits. But analysis of big data and high performance computing that is involved between clouds is challenging.

Following are the applications:

8.1 Data Security:

In order to secure hybrid clouds, companies use special techniques such as authentication, access control policies and encryption in both private clouds and public clouds. These will include the combination of cloud-based security services and managed hosted appliances. Some approaches such as intrusion detection systems and firewalls are always implemented in the hosted environment

8.2 Integration:

One or more private and public clouds integrate to form a hybrid system and it will be more challenging compared integrating the on premises systems.

IV. CONCLUSION

In this paper we have proposed different techniques about data security, privacy and better confidentiality. Deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, for which the duplicate-check tokens of files are generated by the private cloud server with private keys. Using Hybrid cloud architecture, it provides lot of benefits with the use of both public and private. Clouds and adopting deduplication to store data in the cloud will provide us better storage benefits at lower costs.

V. REFERENCES

[1] C. K Huang , L.- F Chien , and Y. -J Oyang , " Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query SessionLogs , " J. Am. Soc. for Information science and Technology, vol. 54, no. 7, pp. 638-649, 2013.

[2] D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 2013.

[3] GNU Libmicrohttpd. http://www.gnu.org/ software/libmicrohttpd.

[4] Ian Clarke, Oskar Sandberg, Brandon Wiley, and Theodore W. Hong"secure clou backup system with assured deletion and version control. In ICDCS, pages 617–624,2013.

[5] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions of Parallel and Distributed Systems, 2013.

[6] Jin Li, Yan Kit Li, XiaoFeng Chen, Patrick P.C.Lee,Wenjing Lou: A Hybrid Cloud Approach for Secure Authorized Deduplication. IEEE transactions on parallel and distributed system Vol:PP NO:99 Year 2014.

[7] John Douceur, Atul Adya, William J. Bolosky, Dan Simon and Marvel, Theimer" Reclaiming space from duplicate files in a serverless distributed file system", 2013. In Workshop on Cryptography and Security in Clouds (WCSC,2011), 2011.

[8] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In ICDCS, pages 617–624, 2002.

[9] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan (2011), Sedic: privacyware data intensive computing on hybrid clouds. In Proceedings of the 18th ACM conference on Computer and communications security, vol. 54, no. 7,2014.

[10] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.

[11] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message - locked encryption and secure deduplication.In EUROCRYPT,pages 296–312, 2014.

[12] M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2013.

[13] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology,22(1):1–61, 2012.

[14] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman - Peleg. Proofs of ownershi in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov,editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.